AFFTC-PA-111038

APPLICATIONS OF STATISTICALLY DEFENSIBLE TEST AND EVALUATION METHODS TO AIRCRAFT PERFORMANCE FLIGHT TESTING

A
F
F
T
C

REAGAN K. WOOLF

AIR FORCE FLIGHT TEST CENTER
EDWARDS AFB, CA

15 December 2011

Approved for public release A: distribution is unlimited.

AIR FORCE FLIGHT TEST CENTER
EDWARDS AIR FORCE BASE, CALIFORNIA
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12-12-2011	Presentation	N/A

4. TITLE AND SUBTITLE Applications of Statistically Defensible Test and Evaluation Methods to Aircraft Performance Flight Testing	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Woolf, Reagan K.	**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES) 773TS/ENFB 412 TW 307 E. Popson Ave. Edwards AFB, CA 93524	**8. PERFORMING ORGANIZATION REPORT NUMBER** AFFTC-PA-111038
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) 773TS/ENFB 412 TW 307 E. Popson Ave. Edwards AFB, CA 93524	**10. SPONSOR/MONITOR'S ACRONYM(S)** N/A **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release A: distribution is unlimited.

13. SUPPLEMENTARY NOTES
CA: Air Force Flight Test Center Edwards AFB CA CC: 012100

14. ABSTRACT
This presentation summarizes seven case studies that illustrate applications of various statistical methods to aircraft performance flight testing. The statistical methods used in the analyses were multi-variable regression analysis, hypothesis testing, and uncertainty analysis. The cases presented were: 1) endurance performance requirement verification, 2) comparison of aerodynamic models, 3) flight test-developed thrust model, 4) climb performance specification compliance, 5) uncertainty analysis of pacer aircraft, 6) evaluation of updated control law, and 7) atmospheric survey for air data calibration. The case studies showed how statistical methods were used to calculate confidence intervals or uncertainty for the final results, detect differences between two aircraft systems with statistical confidence, and reduce the complexity or difficulty of data analysis.

15. SUBJECT TERMS
(T&E) Test and Evaluation regression analysis statistics performance modeling and simulation uncertainty flight test Hypothesis test

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					412 TENG/EN (Tech Pubs)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	None	16	661-277-8615

Standard Form 298 (Rev. 8-98) Prescribed by ANSI Std. Z39.18

Air Force Flight Test Center

AFFTC

War-Winning Capabilities … On Time, On Cost

U.S. AIR FORCE

Applications of Statistically Defensible Test and Evaluation Methods to Aircraft Performance Flight Testing

10 January 2012

Reagan Woolf
reagan.woolf@edwards.af.mil
661-277-3733

Integrity - Service - Excellence

Overview

- Introduction and Motivation
- Aircraft Performance Flight Test Objectives
- Case Studies

Introduction

3

The U.S. Air Force Flight Test Center (AFFTC) is continually seeking ways to improve the planning, execution, analysis, and reporting of developmental test and evaluation programs. Since 2004, statistically defensible methodologies have been studied for their benefits and applicability to the AFFTC mission. Reasons for implementing statistical methods include improving the credibility of flight test results, improving the quality of data used by program offices to make decisions, and optimizing test resources to answer the right questions with the right level of technical risk.

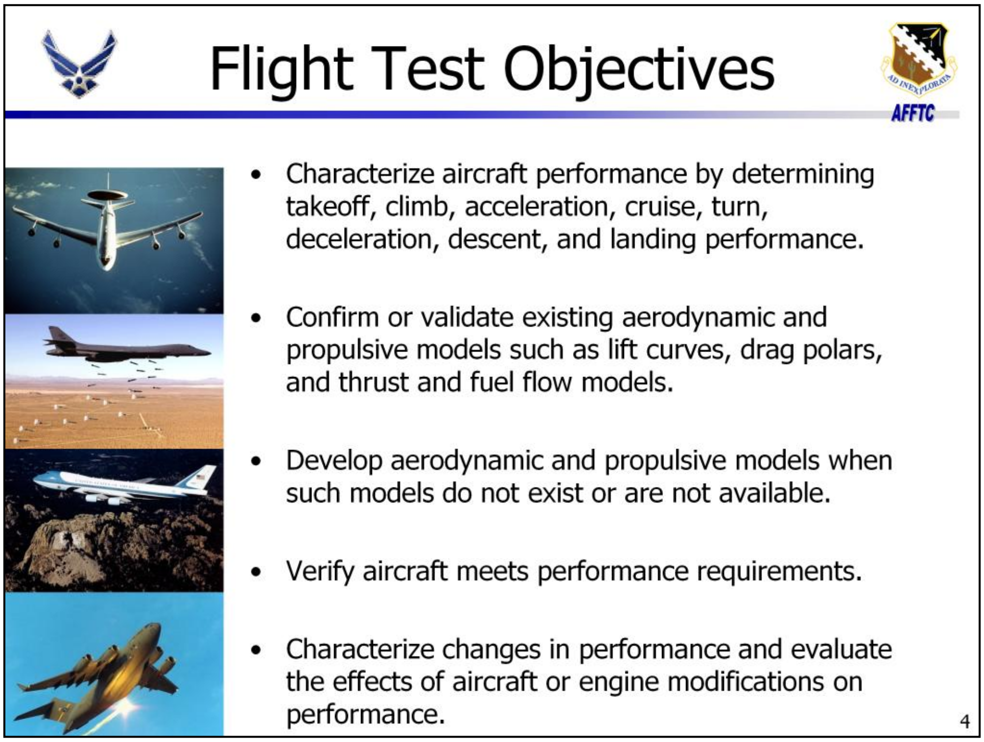
The AFFTC is in the process of training, equipping, and organizing its engineering workforce to implement statistically defensible test and evaluation strategies. Part of this process includes researching and developing new ways to apply statistical analysis techniques to the various engineering disciplines.

We have found that many of our engineers were reluctant to use statistical techniques in their test programs for various reasons, including perceived complexity of the analyses, skepticism over the benefits of statistical analysis, and their limited backgrounds in statistical analysis. Therefore, we started investigating simple methods, such as hypothesis tests or regression analysis, to see which methods added value to our analysis methodologies. Methods that work, and add value, will “sell themselves” to our engineering staff.

This presentation gives seven case studies that illustrate the applications of various statistical methods to aircraft performance flight testing. The methods presented here offered improvements over the “classical” analysis methods. Each of these case studies will compare the “old” analysis methods with the “new.” Some of the improvements included:

1. The ability to calculate confidence intervals, or uncertainty bounds, on the final results
2. The ability to identify differences between systems with statistical confidence
3. The reduction in complexity or difficulty of the analysis.

While the use of these statistical methods added value to the data analysis methods, it did not increase the efficiency of test execution nor did it reduce the overall amount of testing required.



The objectives of performance testing are to:

•Characterize aircraft performance by determining takeoff, climb, acceleration, cruise, turn, deceleration, descent, and landing performance.

•Confirm or validate existing aerodynamic and propulsive models such as lift curves, drag polars, and thrust and fuel flow models.

•Develop aerodynamic and propulsive models when such models do not exist or are not available.

•Verify aircraft meets performance requirements.

•Characterize changes in performance and evaluate the effects of aircraft or engine modifications on performance.

Several statistical methods that have proven useful in support of these objectives are multi-variable regression analysis, hypothesis testing, and uncertainty analysis. Regression analysis has been used to develop models of multi-variable data sets, and hypothesis testing has been used to determine which coefficients in the regression models were significant. Hypothesis testing has also been used to compare the means and variances of two samples of data. Uncertainty analysis has been used to estimate the systematic and random components of data uncertainty.

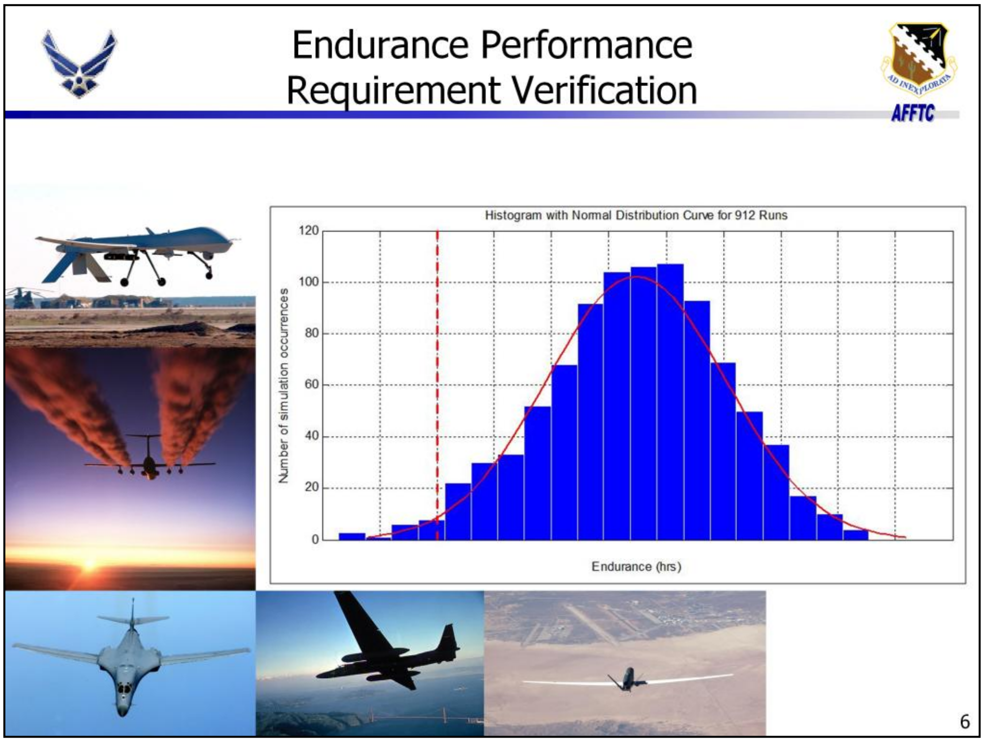
Case Studies

AFFTC

- Endurance Performance Requirement Verification
- Comparison of Aerodynamic Models
- Flight Test-Developed Thrust Model
- Climb Performance Specification Compliance
- Uncertainty Analysis of Pacer Aircraft
- Evaluation of Updated Control Law
- Atmospheric Survey for Air Data Calibration

5

These seven case studies illustrate the applications of multi-variable regression analysis, hypothesis testing, and uncertainty analysis to problems in aircraft performance flight testing.



Test Objective:

- Determine if an aircraft has a minimum total endurance of X hours

Test Approach:

- Collect flight data regarding drag and lift coefficients, fuel flow, and air data
- Validate aerodynamic and propulsive models using flight test data
- Use aircraft performance simulation to predict endurance
- Compare predicted endurance to minimum requirement

Analysis Approach:

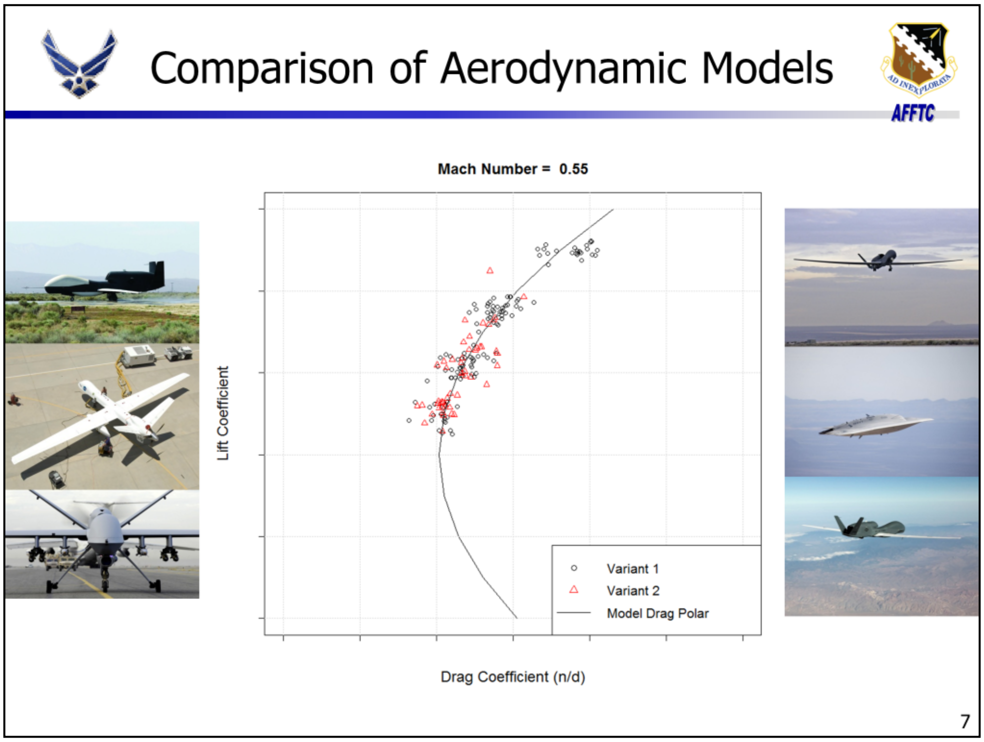
- OLD
  - Estimate uncertainties with scatter bands on final results.
- NEW
  - Estimate uncertainties associated with weight, outside air temperature, thrust, fuel flow, drag coefficient, and calibrated airspeed
  - Use Monte-Carlo analysis to propagate uncertainties and calculate expected endurance

Results:

- Aircraft met requirement
- Lower 95 percent confidence bound exceeded requirement.

Benefits:

- Allowed us to propagate elemental uncertainties through complicated non-linear performance simulation to estimate overall uncertainty in endurance.
- Presented final results with 95-percent confidence interval, which increased credibility of conclusion and quantified technical risks due to uncertainties in flight test data.



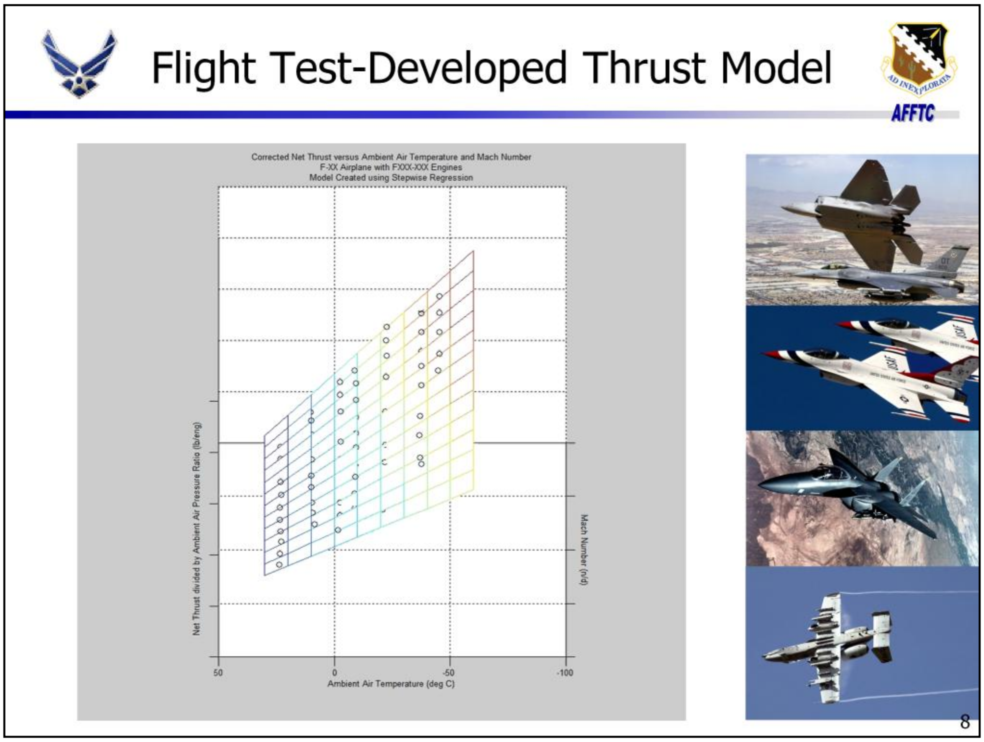
Test Objective:
- Determine if the aerodynamic differences between two aircraft variants were large enough to warrant different flight manual performance charts

Test Approach:
- Collect flight test lift and drag data during dedicated max power climbs, partial power cruise, and idle power descents
- Collect “target of opportunity” data during other test events to increase size of lift and drag coefficient data base

Analysis Approach:
- OLD
    - Inspect residuals (differences between data and model fit)
    - Two distinct levels of the residuals implied that the two aircraft variants were different
- NEW
    - Use multi-variable regression and **parallel lines tests** to compare the drag polars of the two variants
    - Included “aircraft variant” as a regressor in the least squares model.
    - Tested the regressor to see if its coefficient was significantly different than zero. (Null hypothesis was regression coefficient was zero. P-value greater than 0.05 implied coefficient was not significant.)

Results:
- No significant differences in drag
- P-value of “aircraft variant” term was 0.73, which was greater than level of significance of 0.05. Failed to reject the null hypothesis, which implied that the “aircraft variant” coefficient was zero.

Benefits:
- Regression coefficient of the “aircraft variant” term provided a measure of the difference in drag between the two variants.
- Analysis method eliminated some of the subjectivity associated with judging the differences between the two aircraft variants.
- Able to detect small differences in drag due to large quantity of data.
- This analysis was successful because of the large quantity of data, most of which was not pre-planned. Around 75 percent of the data came from targets of opportunity. Other programs with less data available will not be able to detect such small differences.



Test Objective:

- Develop flight test-based, empirical, installed net thrust model

Test Approach:

- Measure excess thrust during MIL power level accelerations and sawtooth climbs
- Assume correct aero drag model. Calculate net thrust:
  Net Thrust = Excess Thrust - Drag

Analysis Approach:

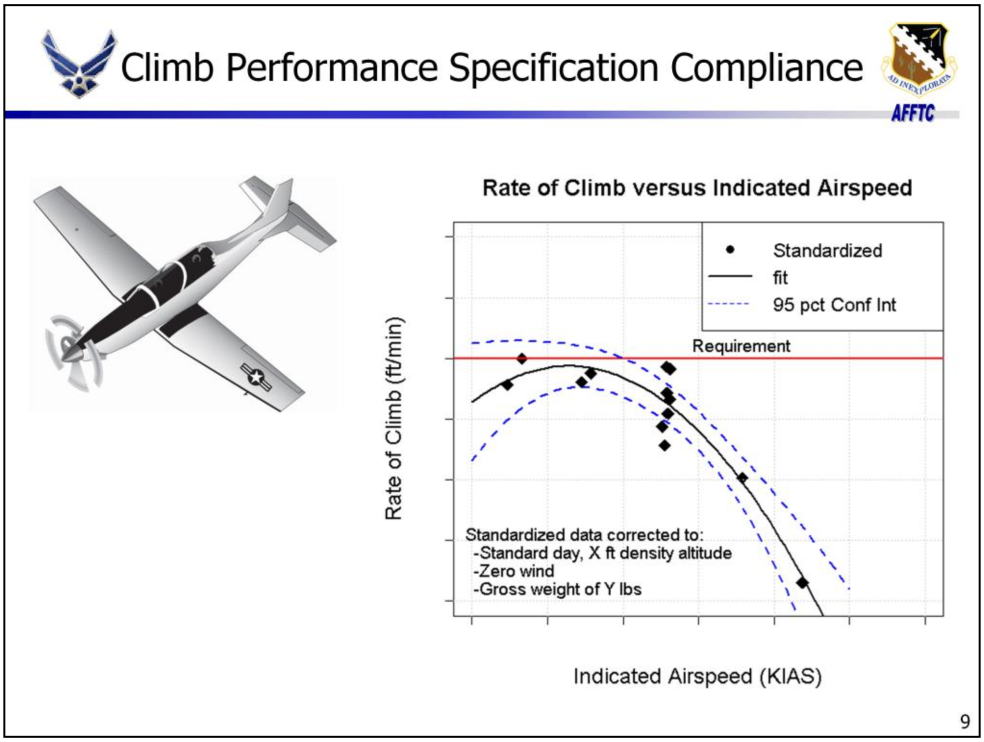
- OLD – Piecewise linear regression
  - Divide net thrust data into Mach number bins. For each bin, plot net thrust divided by ambient air pressure ratio versus ambient air temperature
  - Fit line to each bin
  - Force slopes and intercepts of each line to follow smooth curve in Mach number
- NEW – Multi-variable, stepwise linear regression
  - Analyst provides candidate regressors and interactively develops regression model
  - Stepwise regression tool includes statistical parameters to help the analyst:
    - choose which regressors to include in model (squared partial correlation)
    - choose which regressors to discard from the model (F ratio)
    - evaluate the goodness of the model (R squared, predicted squared error, and percent fit error)
  - Candidate regressors: M, T, M^2, T^2, M^3, T^3, M*T

Results:

- Final model: Fn/delta = beta0 + beta1*M + beta2*T + beta3*M*T
- R squared = 0.9893
- “Full” model resulted in slightly higher R squared, but final model almost as good and much simpler (parsimony principle)

Benefits:

- Reduced the complexity of the analysis. The OLD analysis method (piecewise regression) was very time consuming (sometimes took days). The NEW multi-variable, stepwise regression method could be accomplished in hours.



Test Objective:

- Demonstrate the aircraft has ability to climb at least **Z** feet per minute at **X** feet density altitude and **Y** pounds gross weight

Test Approach:

- Perform sawtooth climbs at 5 airspeeds, with extra replicates near predicted airspeed for maximum rate of climb
- Corrected test data to standard conditions

Analysis Approach:

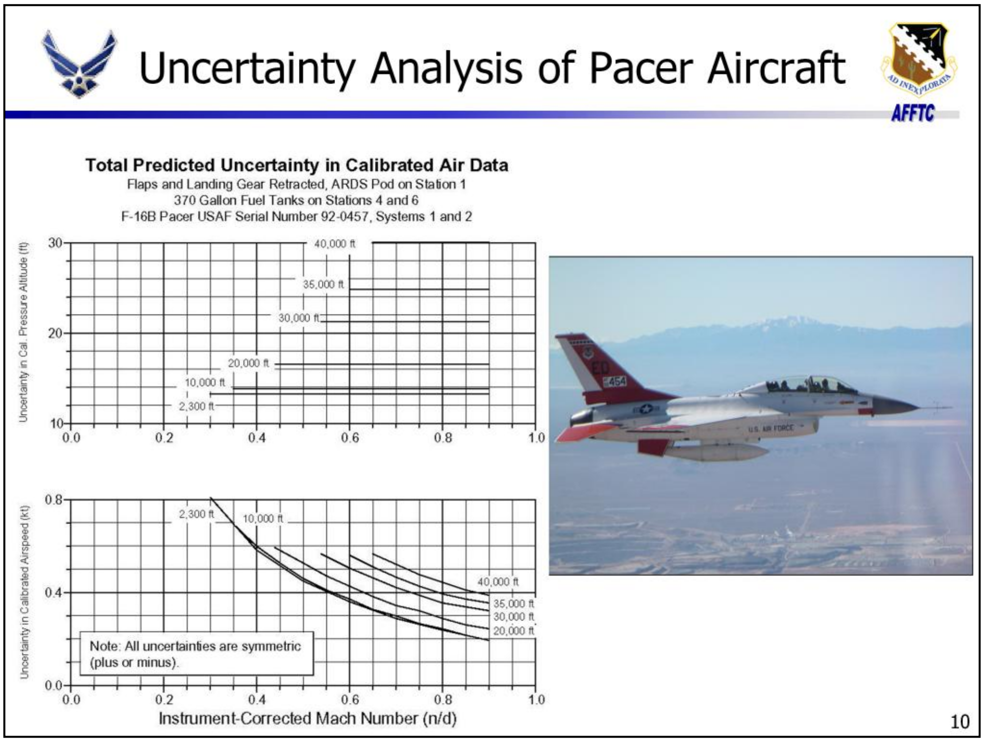
- OLD
    - Fit regression model to data
    - Compare fit and data scatter to requirement line
- NEW
    - Fit regression model to test data and generate 95-percent confidence intervals

Results:

Test data and regression model were below requirement. However, upper confidence interval exceeded requirement. Therefore, concluded the aircraft performance met requirement, but was borderline.

Benefits:

- Provided a 95-percent confidence interval about the line fit, which served as a measure of the uncertainty in the flight test results.
- Eliminated some of subjectivity associated with judging if the rate of climb met, or came close enough, to the requirement even though all of the data points were below the requirement line.



Test Objective:

- Determine uncertainty in calibrated air data for the AFFTC F-16 pacer aircraft

Test Approach:

- Calibrate pacer aircraft using standard methods

Analysis Approach:

- OLD
  - Estimate uncertainty using scatter bands on calibration curves
- NEW
  - Perform uncertainty analysis (per ASME test uncertainty standard) on entire pacer calibration process
  - Trace uncertainties of all truth source and pacer instruments back to lab standards (which were traceable to NIST)
  - Propagate uncertainties to final calibrated air data parameters

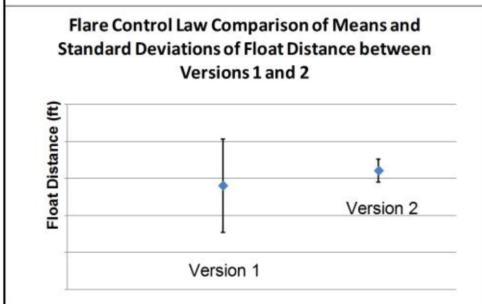
Results:

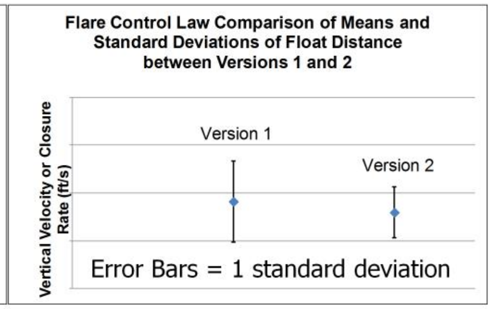
Pacer uncertainties are 30 feet and 0.8 knots at worst case.

Benefits:

- Communicated uncertainties of pacer data products to customers
- Uncertainties were estimated using industry-accepted practices (ASME test uncertainty standard).

Updated Control Law





Measurement	Parameter	Test	p value	Conclusion
Float Distance	mean	*t*-test	0.11	fail to reject H_0
	variance	*F*-test	0.0002	reject H_0
Vertical Velocity at Touchdown	mean	*t*-test	0.44	fail to reject H_0
	variance	*F*-test	0.09	fail to reject H_0

11

Test Objective:

- Compare landings with new flare control law to landings with old flare control law

Test Approach:

- Perform landings on flat and sloped runways at light and heavy aircraft gross weights
  - Slope: 3 levels (flat, uphill, downhill)
  - Weight: 2 levels (heavy and light)
- Compare "float distance," which was the distance between the touchdown aim point and the actual touchdown point
- Compare vertical velocities at touchdown

Analysis Approach:

- OLD
  - Visually compare means and data scatter
- NEW
  - Perform statistical testing
  - Use t-test to compare means (H0: means are equal, H1: means are not equal)
  - Use F-test to compare variances (H0: variances are equal, H1: variances are not equal)

Results

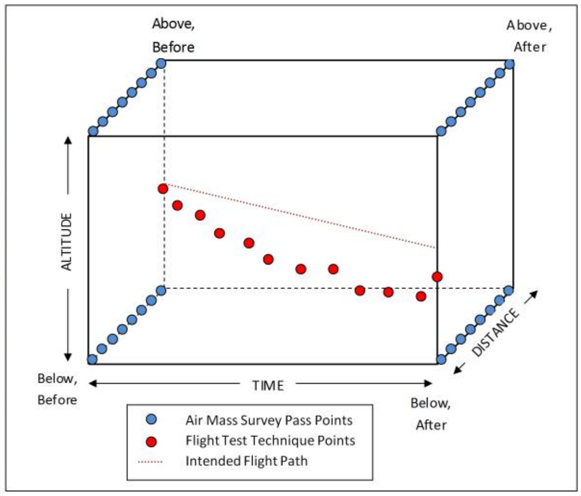
- Mean float distances and vertical velocities were similar (fail to reject H0)
- Variances of vertical velocities at touchdown were similar (fail to reject H0)
- Variances of float distances were DIFFERENT (reject H0). The new law had much smaller variance, meaning the aircraft touchdown points had less dispersion.
- Conclusion: new control law was as good or better than old control law

Benefits:

- Enabled the analyst to detect differences with statistical confidence and power.
- Eliminated the subjectivity associated with comparing the results from the two flare control laws

Atmospheric Survey for Air Data Calibration

AFFTC



- **Calibration Equation:** $H_c = -84 + 0.9606 H_{GPS} - 0.0039t + 0.1979d$

12

Test Objective:
- Determine the static source error corrections for a flight test noseboom

Test Approach:
- Survey the atmosphere in the test range using the "self survey" method
- Fly pairs of high and low passes through range, before and after flying a level acceleration/deceleration
- Level accel/decel sweeps through full range of Mach numbers

Analysis Approach:
- OLD
  - Use single weather balloon to survey column of air. Plot atmospheric pressure versus geometric (GPS) altitude
  - Assume survey of column of air was valid throughout test range
  - Use survey as truth source for ambient air pressure versus geometric altitude
  - Track aircraft using differential GPS and compare to survey data
- NEW
  - Don't assume survey from single balloon is valid. Develop model that accounts for changes in truth source pressure with variations in geography and time.
  - Use multi-variable regression and data from survey passes to model changes in pressure with GPS altitude, distance from a reference point, and time
  - Check data for serial correlation. Sample data by appropriate number of lags to remove serial correlation before creating linear model.
  - Test each regressor to determine significance. For example, were "time" or "distance squared" significant terms in the model?

Results:
- Model successfully used in calibration.
- Model results were similar to results obtained using other methods.

Benefits:
- New flight test technique enabled nearly self-contained calibration of Pitot-static system without the need for external support assets (such as other pacers, multiple weather balloons, radar tracking)
- New analysis method (multi-variable regression) didn't require the (often bad) assumption that a single weather balloon was sufficient to survey entire airspace in test range.

Conclusions

- Express uncertainty
- Detect differences
- Reduce complexity

13

Statistically defensible test and evaluation methods have been successfully applied to several problems in aircraft performance flight testing. These methods offered improvements over existing analysis methods, such as:

1. Conclusions were expressed with confidence intervals or uncertainty bounds, as in the performance verification cases and the pacer uncertainty analysis.

2. Differences (or no differences) between systems were determined with statistical confidence, as in the case of identifying the aerodynamic differences between two aircraft variants.

3. The reduction in complexity or difficulty of some analyses, as in the case of developing multi-variable regression models of thrust or atmospheric pressure.

The application of these methods provided tangible benefits, which means that we should have no trouble convincing our engineering staff to adopt and expand these methods to new applications.

However, these case studies showed that although the use of statistics added value to the data analysis methods, its use did not necessarily improve test efficiency or reduce the overall number of test points. In a few cases, such as the comparison of aerodynamic models of two aircraft variants and the thrust model development, the very large quantities of data that were available led to the successful application of the statistical methods. These methods may not be as successful on other test programs for which less data are available.

Questions?

14